**Sarah E. Cornwell**
**The University of Western Ontario, London, Ontario, Canada**

# Language Classification in *The Ethnologue* and its Consequences (Paper)

**Abstract:**

*The Ethnologue* is a widely used classificatory standard for the world's 7000+ natural languages. However, the motives and processes used by *The Ethnologue*'s governing body, SIL International, have come under criticism by linguists. This paper investigates how *The Ethnologue* answers the question "What is a language?" through the theoretical lens presented by Bowker and Star in *Sorting Things Out* (1999) and presents some consequences of those classificatory decisions.

## 1. Introduction

Language is central to human culture and experience. It is the central core of our communicative capacity, running through all media of communication. Due to this essential role, there is a need to classify and count languages and their speakers[1]. Governments need to communicate with citizens, libraries need to catalogue books by language, and search engines need to return webpages in the appropriate language for each searcher. For these reasons among others, there is a strong institutional need for a standardized classification structure and labelling system for languages. The most widespread current classificatory infrastructure for languages is based on *The Ethnologue*. This essay will critically evaluate *The Ethnologue* using the Foucauldian theory developed for investigating classification structures by Bowker and Star (1999) in *Sorting Things Out: Classification and its Consequences*.

## 2. *The Ethnologue*

*The Ethnologue* is a catalogue of all the world's languages. First compiled in 1951, *The Ethnologue* is currently in its 20th edition and contains descriptions of 7099 living languages (2017). Since 1997, *The Ethnologue* has been available freely online and widely accessible (Simons & Fennig, 2017). In addition to naming languages, *The Ethnologue* also classifies them into language families, rates their health using the Expanded Graded Intergenerational Disruption Scale (EGIDS), and lists basic typological elements, countries in which the language is spoken, population of speakers, and which, if any, writing system is used (Simons & Fennig, 2017). Due to its comprehensiveness and accessibility, *The Ethnologue* has become the primary text used for understanding the scope and health of world languages.

---

[1] I use "to speak" and "speaker" in this essay for simplicity, but I am not excluding signed languages. Spoken languages make up the majority of human languages, but signed languages include the same complex features, and are equally "language" in every understanding of the word.

*The Ethnologue* is published by SIL International, formerly known as the Summer Institute of Linguistics. SIL is not an academic linguistics organization, but is a not-for-profit religious group whose linguistic work is in support of their main aim of bible translation (Campbell & Grondona, 2008). As such, *The Ethnologue* has a strange place in linguistic scholarship. Google Scholar records around 8000 citations of various editions of *The Ethnologue* since 1996, despite receiving strong criticism from many linguists. In their review of the 15th edition, Campbell and Grondona (2008) criticize *The Ethnologue*'s consistency, accuracy, age of references, and stronger focus on common religious affiliation than basic typological information. In addition, Campbell and Grondona (2008) have serious issues with the classification of languages, noting that *The Ethnologue* frequently lists more languages per family than specialists do, that it includes a number of disproven or unsupported language families, and that it often confuses the labeling of isolate and unclassified languages. Despite this harsh criticism, they end their review with the note that *The Ethnologue* "is truly excellent, highly valuable, and the very best book of its sort available" (Campbell & Grondona, 2008, p. 640). Hammarström (2007, pp. 14-15) concurs, offering similar criticism before finally noting that *The Ethnologue* "continues to be the best single source on the living languages of the world, in spite of its bad sides".

The international community seems to agree with linguists' assessments. In 2007, the 15th edition of the Ethnologue was used to create the first version of ISO 639-3, and SIL International was appointed as the registration authority for this standard (SIL International, 2017). ISO 639-3 is the International Organization for Standardization's (ISO) comprehensive listing of all languages using a 3-letter code (ISO, 2017). Due to the wide use of ISO standards, SIL's position as the only registration authority gives SIL monopolistic control over the world's understanding of what is and is not a human language. The ISO standards are used globally by libraries and publishers who use the Library of Congress system, as well by major players in technology. ISO 639-3 is the basis for the Internet Engineering Task Force's (IETF) standardized language tag system (used as the HTML5 standard for labelling websites) (Phillips & Davis, 2009), as well as the system used in Microsoft Windows 10, and by the Wikimedia Foundation (Meta Contributors, 2012). This monopolistic control has been criticized by linguist groups, especially due to the potential for misuse of data concerning vulnerable indigenous populations by missionary groups such as SIL (Epps et al., 2006). Other concerns were raised about errors in *The Ethnologue*'s data, pejorative designations used for some language names, and the feasibility of creating permanent codes for languages which are inherently impermanent (Morey, Post, & Friedman, 2013). However, finding no alternative of acceptable quality, most groups, (including the Society for the Study of the Indigenous Languages of the Americas (SSILA) which had been one of the strongest critics) decided the need for a high quality standard outweighed concerns about SIL (Golla & Scott, 2006). Understanding the issues with classification in *The Ethnologue* becomes more urgent when its central position of control is well understood.

Per Bowker and Star (1999), these applications of *The Ethnologue* make it not just a classification, but also a standard. As they are under tension from stakeholders from many communities, standards tend to be stable and resist changes (Bowker & Star, 1999, pp. 13-14). Straddling these communities of practice also makes *The Ethnologue* a boundary object: "inhabit[ing] several communities of practice and satisfy[ing] the informational requirements of each of them" (Bowker & Star, 1999, p. 16).

### 3.  Classificatory Consequences

Classifications in *The Ethnologue* are done according to three criteria set in ISO 639-3 (Simons & Fennig, 2017). The foremost criterion is mutual intelligibility. Where intelligibility is marginal, the second criterion allows for languages with a common literature or ethnocultural understanding to be grouped together (Simons & Fennig, 2017). The final criterion states that where ethnic groups have distinct identities, even if they can communicate well, their languages could be listed separately. There are two serious problems with these criteria.

First, claiming ISO 639-3 as the source of language classification criteria when *The Ethnologue* was the source (and while SIL is the registration authority) for ISO 639-3 is strangely obscuring and circular. SIL is transparent about being the source for ISO 639-3 (see: SIL International, 2017), so why be evasive? *The Ethnologue* seems to be borrowing institutional cultural capital from the ISO to bolster and legitimate its findings. As a non-academic, non-governmental institution, SIL has little institutional cultural capital of its own (Bourdieu, 2011/1986). Referencing the ISO in this way leads readers to assume that the language classifications have been vetted or otherwise approved by a large, secular, international, semi-governmental technical organization. The SIL gains capital through this assumption, though it could hardly be further from the truth. In fact, SIL has complete control over the ISO's language code standard through its role as the registration authority. The circularity here allows SIL to borrow cultural capital from the ISO while maintaining control over the standard.

The second major concern is that the second and third criteria effectively overrule the first. The designations of language and dialect are fluid. As Max Wienberg famously said: "A language is a dialect with an army and a navy" (Zenderland, 2014). This definition is true in *The Ethnologue* as well. The designation of "language" is not determined by any collection of internal features[2], but by political and ethnocultural factors. The criteria used in *The Ethnologue* to delineate languages are not Aristotelian binaries but are entirely dependent on their similarity to prototypical ideas of languages and their speakers. Specifically, the idea that language groups are composed of a relatively ethnically homogenous group of speakers. While languages can be socially defined, they do not align perfectly with ethnocultural boundaries. Classification in *The Ethnologue* lends support to the assertion in Bowker and Star (1999, p. 64) that classification

---

[2] A system could instead use the presence or absence of, for example, phonological features like clicks or nasal vowels, syntactic features like word order, or other grammatical features like case or grammatical gender. See the World Atlas of Language Structures at www.wals.info for an example.

systems "reflect the conflicting, contradictory motives of the sociotechnical situations that gave rise to them". As such, the technical classification of languages might have to "grow out of and answer to our common sense, socially comfortable classifications" (Bowker & Star, 1999, p. 67). The classifications within *The Ethnologue* must be understood by, and reflect the beliefs of, the multiple stakeholders who access them. As a missionary organization, SIL has a vested interest in maintaining cordial relationships with the states and ethnic populations that they do their work with.

## 4. Colonization Narratives in *The Ethnologue*

"Ethnic and racial classifications are a mechanism of power that establishes inequality in a society as a result of the organization of diverse groups into a certain number of categories based on their allegedly cultural and biological features […] Linguistic classifications are not an exception to this rule" (Mamontova, 2016, p. 48). Or as Bowker and Star write "classification systems are necessarily imperialistic" (1999, p. 276). This is also true of language classifications done in *The Ethnologue*. The answer to the question of "is this a language?" can have real political, social, and economic consequences for its speakers.

The majority of languages in the world are spoken by a tiny minority of people. This can be illustrated with the fact that the 300 largest languages (out of a total of over 7000 living languages) are spoken by over 90% of the world's population. This means that less than 10% of the world's population are our only stewards for over 95% of our linguistic diversity (Simons & Fennig, 2017)[3]. The way that we treat the speakers of these languages determines the future of not only their language, but of all the cultural knowledge and understanding that is intertwined within the language.

Orok is an indigenous language of the Manchu-Tungusic family, spoken in Siberian Russia (Simons & Fennig, 2017). *The Ethnologue* reports a total of around 300 ethnic Orok peoples, of whom only 50 are fluent in the Orok language (Simons & Fennig, 2017). Orok has two main dialects with little mutual intelligibility, and was officially considered a dialect of Nanai by Russia (Simons & Fennig, 2017). Mamontova (2016) notes that the ethnic community in Northern East Asia is heterogeneous, making the labelling of ethnic and language groups a "difficult and rather unnatural task" (p.53). From the Soviet point of view, "only an ethnic group whose dialect occupied both a linguistically and geographically central position could be treated as a nation"(Mamontova, 2016, p. 54). As Orok was considered a dialect, the Soviet government began imposing literacy training for Orok peoples as a part of language standardization efforts in the region (Mamontova, 2016). The Nepa dialect of Evenki was taught because it was considered more "literary" and developed by colonial Soviets (Mamontova, 2016). This was effectively a colonization project, intended to remove cultural differences between the different groups of Tungus-speaking peoples and draw them closer to the culture of the western parts of Russia

---

[3] Specifically Table 2 at https://www.ethnologue.com/statistics/size

(Mamontova, 2016). It had the effect of relabeling indigenous peoples within broader ethnicities and created new social hierarchies of indigenous groups (Mamontova, 2016).  Those who speak the chosen "standard" dialect hold more social capital; their identity and rights as indigenous peoples of Russia fit them more comfortably. Those who speak another dialect have had that capital stripped – they must become comfortable with their new ethnic label or abandon their special rights as indigenous peoples; their people's name erased (Mamontova, 2016). This erasure is an example of how, as Bowker and Star (1999, p. 6) write, "for any individual, group or situation, classifications and standards give advantage or they give suffering". To simplify categories within language classification standards there is no "not otherwise specified (NOS)" language category (unlike the NOS causes of death categories Bowker and Star (1999) found in the ICD). Every language is classified, and smaller varieties may be consumed against their will by colonizing classificatory forces.

The problems raised by language classification in *The Ethnologue* are not easy to solve. Languages are inherently difficult to classify. Languages are constantly changing and perfectly heterogeneous: no two speakers are exactly alike in their understanding. Philosophers of language have found language so difficult to define that some have argued languages do not actually exist (See Stainton (2016) for a summary of arguments made). Leaving languages unclassified is an equally impossible task. Technologies of the modern world require language classification infrastructures to function. And despite its issues, I must agree with previous reviewers who have found *The Ethnologue* to be both flawed but the current best option.

**Reference List:**

Bourdieu, P. (2011/1986). The forms of capital (R. Nice, Trans.). In I. Szeman & T. Kaposy (Eds.), *Cultural theory: An anthology* (pp. 81-93). UK: John Wiley & Sons Ltd.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences* (paperback ed.). USA: MIT Press.

Campbell, L., & Grondona, V. (2008). Review: Ethnologue: Languages of the world by Raymond G. Gordon. [Ethnologue: Languages of the World]. *Language, 84*(3), 636-641.

Campbell, L., & Poser, W. J. (2008). *Language classification: History and method*. New York, USA: Cambridge University Press.

Cerruti, M., & Regis, R. (2014). Standardization patterns and dialect/standard convergence: A northwestern Italian perspective. *Language in Society, 43*(01), 83-111. doi:10.1017/S0047404513000882

Epps, P., Mateo-Toledo, B. a., Hallett, T., England, N., Cruz, H., Rauschuber, B., . . . Villard, S. (2006). In opposition to adopting Ethnologue's language codes for ISO 639-3 (pp. 2-3). The Society for the Study of the Indigenous Languages of the Americas (SSILA) Bulletin: SSILA.

Gao, X. (2015). The Ideological Framing of 'Dialect': An analysis of mainland china's state media coverage of 'dialect crisis' (2002-2012). *Journal of Multilingual and Multicultural Development, 36*(5), 468-482. doi:10.1080/01434632.2014.943234

Golla, V., & Scott, D. (2006). SSILA statement on iso 639-3 language codes (pp. 1-5). The Society for the Study of the Indigenous Languages of the Americas (SSILA) Bulletin: SSILA.

Hammarström, H. (2007). *Handbook of descriptive language knowledge: A full-scale reference guide for typologists*. Muenchen: LINCOM GmbH.

Hinrichs, E., Gerdemann, D., & Nerbonne, J. (n.d.). *Measuring linguistic unity and diversity in Europe*. Proposal.  Retrieved from https://www.researchgate.net/publication/241125011_Measuring_linguistic_unity_and_diversity_in_Europe

ISO. (2017). Iso 639 language codes. Retrieved from https://www.iso.org/iso-639-language-codes.html

Jensen, J. B. (1989). On the mutual intelligibility of Spanish and Portuguese. *Hispania, 72*(4), 848-852. doi:10.2307/343562

Laharnar, N. L. (2003). The Balkans: A mosaic. *Ideas, 1*(1), 21-38.

Mair, V. (2011, February 14, 2011). Cantonese Blackberry ad.  Retrieved from http://languagelog.ldc.upenn.edu/nll/?p=2970

Mamontova, N. (2016). Language as a mechanism of systemic foundation: Tungus-speaking groups in the far east. *Asian Ethnicity, 17*(1), 48-66. doi:10.1080/14631369.2015.1086089

Meta Contributors. (2012, March 19). Language tagging. Retrieved from https://meta.wikimedia.org/w/index.php?title=Language_tagging&oldid=3579530

Morey, S., Post, M. W., & Friedman, V. A. (2013). *The language codes of ISO 639: A premature, ultimately unobtainable, and possibly damaging standardization*. La Trobe University.  Retrieved from https://ses.library.usyd.edu.au/bitstream/2123/9838/1/009 MoreyPostFriedman.pdf

Phillips, A., & Davis, M. (2009). Tags for identifying languages. Retrieved from http://www.rfc-editor.org/rfc/rfc5646.txt

SIL International. (2017). ISO 639-3 Registration Authority - SIL International. Retrieved from http://www-01.sil.org/iso639-3/default.asp

Simons, G. F., & Fennig, C. D. (Eds.). (2017). *Ethnologue: Languages of the world* (20 ed.). Dallas, Texas, USA: SIL International.

Stainton, R. J. (2016). A Deranged Argument against Public Languages. *Inquiry, 59*(1), 6-32. doi:10.1080/0020174X.2015.1115279

Tang, C. (2009). *Mutual intelligibility of Chinese dialects: An experimental approach.* (PhD Dissertation), Universiteit Leiden, Utrecht, Netherlands. (LOT;228)

Williams, J. M. (1975). *Origins of the English language: A social and linguistic history*. New York, USA: The Free Press.

Zenderland, L. (2014). Social science as a "weapon of the weak": Max Weinreich, the Yiddish scientific institute, and the study of culture, personality, and prejudice. *Isis, 104*(4), 742-772. doi:10.1086/674942